Good Practice in Using Statistics in Statistics Education Research

Neville Davies and Gemma Parkinson
Royal Statistical Society Centre for Statistical Education, Plymouth University, Plymouth, UK, neville.davies@rsscse.org.uk

In 2007 the American Statistical Association (ASA) published a report 'Using Statistics Effectively in Mathematics Education Research.' In October 2012 a workshop was organized in the UK to discuss how the mathematics education community could make use of the report's findings and recommendations. In this paper we briefly reflect on the ASA report and the discussions. We then suggest that the statistics and statistics education community should use both the ASA report's recommendations and other developments, such as the work of the National Foundation for Educational Research (and others) in the UK on randomized controlled trials in education research. We use two examples of data visualization tools and consider influential factors for carrying out interventions with them. Allowing for these factors leads to possible engagement with, and fostering cooperation between, the statistics and statistics education communities. In general such cooperation could, in turn, produce a more convincing evidence-base for improving statistics education.

Key Words: interventions; randomized controlled trials; cooperation; data visualization

1. Introduction

For many years qualitative research has dominated the practice and publication of educational interventions. Indeed, quantitative research approaches, such as those used in experimental design, are rarely used in either generic or subject-specific educational research. When statistical methods are used, often we have noticed a poor standard of application. For example, papers in recent issues of the British Educational Research Journal reveal fundamental statistical flaws in the design and implementation of quantitative research across a wide range of subjects. The most basic flaws appear to be that researchers fail to define their target populations, even though they happily carry out hypothesis tests on 'samples' of learners, sometimes from a single classroom in a conveniently located school.

Interventions are common when new and existing methods of teaching or 'doing' a subject need to be compared. For example 'doing' medicine often involves a need to know the effect of a drug when administered to a population of individuals. The penalty for failure or getting the drug wrong in some way can be catastrophic for the people who took part in the original research and for the target population.

For that reason a gold standard, randomised controlled trials (RCTs), is applied to the testing and evaluation of drugs by statisticians and medical researchers. This involves careful design that uses randomisation in such a way that fair comparisons can be made, ensuring groups of individuals are equivalent in terms of factors that might bias conclusions about the drugs.

2. Education Interventions

Education interventions are important activities that can help both teachers and learners. Indeed such interventions can also influence policy of governments in their desire to improve education for all. However, the penalty of failure with an educational intervention is unlikely to be a matter of life or death, as in trials for medical research. The use of RCTs in educational research in any subject is rare. Some of the reasons are discussed in Hutchison and Styles (2010), which is an excellent guide to running RCTs for education researchers. It presents the advantages of using RCTs, but also discusses the common objections to them including: ethical issues; generalizability; inability to carry out 'blind' interventions; the complexity of educating and being educated; and knowing *how* interventions work. In the last case the authors argue for carrying out qualitative and quantitative research simultaneously.

Hutchison and Styles highlight differences and similarities between educational and medical interventions and make recommendations with regard to sample size, testing before interventions, data structure problems, measurement and treatment of drop-outs. The authors cogently argue that RCTs should be adopted as the gold standard for some evidence-based educational research and practice. Statistics education researchers may find some of the proposals very useful.

Other influential British authors have advocated an increased use of RCTs across many subjects that affect the public, including education. In a broadcast on BBC Radio 4 in January 2013, the author of the best-selling book *Bad Science*, Ben Goldacre, hosted a discussion about the merits of RCTs in policy design, implementation and evaluation. The same author, (Goldacre, 2013) produced a short article about building evidence into education practice. The key recommendations were: research on what works best should be a routine part of life in education; teachers should be empowered to participate in research; myths about randomized trials in education should be addressed, removing barriers to research. The article provoked diametrically opposed views on the comment blog. Recommending universal use of RCTs is controversial, especially as some of the theoretical assumptions underpinning RCTs are not satisfied in educational research.

3. Using Statistics in Mathematics Education Research

The 2007 ASA report aimed to develop a stronger foundation of research in mathematics education that is scientific, cumulative, interconnected and weaved with teaching practice. It considered four research practices that deserve special consideration: measurement; RCTs; experimental versus observational research; and intervention gain scores.

At a workshop organised for mathematics education researchers in the UK in 2012, the ASA report was discussed and a number of items discussed. These included: whether or not mathematics education research is influential to policy making; the connections between mathematics education research and teaching practice; and the extent that the findings from mathematics education research are generalisable.

Several mathematics colleagues were sceptical about the approach advocated in the ASA report, with one of them finding the tone of the report unacceptable. 'Mathematics Education research is different' he commented. Others argued that the mathematics education community should take heed of many of the faults identified by the authors of the ASA report; some that careful account should be taken of the data structures that surround intervention measures. For example, one very effective way to do this is through multilevel modeling (MLM) (Goldstein, 2010). This can provide a statistically sound way to model interventions in mathematics (and other subject-specific) education research, especially when factors are nested. It was agreed that more people should adopt sound statistical modeling approaches typified by MLM to measure the effect of interventions. However, some of the studies reported by the ASA that used RCTs were flawed in assumptions and design.

The meeting prompted us to consider whether *statistics* education researchers could benefit from the kind of discussions that took place with mathematics education researchers and whether designed experiments or modelling, for example RCTs and MLM, could be utilised in assessing the efficacy of statistics education interventions.

4. Using Statistics in Statistics Education Research

Statisticians, government statistical organisations, users and producers of statistics and teachers of statistics in schools and higher education are increasingly using *data visualisation* (DV) to communicate and tell data 'stories', often about very large and complicated data structures. DV usually involves a range of static and dynamic images with interaction allowed from the user.

Rather than revisit interventions that others have used for the improvement of teaching and learning traditional programmes of study in statistics, we discuss what needs to be considered before carrying out interventions in statistics education, such as a DV intervention. Investigating the effectiveness of DV tools in teaching and learning is part of a new research project being run in the UK.

Because visual interpretation can be a personal and subject activity, we consider how a DV tool might be created from design through to its final use in informing or educating a user. The constructor of a DV tool will have in mind the conclusions that are *intended* to be drawn from the designed dynamic display of data. However, these may not coincide with the actual conclusions that are drawn by a user of the display. Similarly, teachers that use DV tools to educate their students may not teach what the constructor intended and, even more importantly, the students may not learn what the teacher intended – there could be further ramifications if assessments are constructed using the DV tools. These comprise pre-intervention factors.

As well as design and fit-for-purpose issues, the amount of information that can be gleaned and retained from a highly visual approach to getting information from data, leads to the need to consider the following factors in administering interventions:

- the optimal balance of static versus dynamic images for DV;
- the use of different colours in the images to provide the greatest accessibility;
- the features that characterise good and bad DV;
- the cognitive skills needed to get trustworthy information using DV;
- age-, gender- or subject-specific-dependency on the amount of information that can be gleaned from DV;
- the educational backgrounds of users of DV;
- the educational value of, and what can be learned from DV, and the time taken to learn how to use the DV tool;
- the effectiveness of keeping the user engaged to explore other problems.

The design and post-use factors demonstrate that in trying to assess the value of using DV in teaching, learning, assessment or providing the public with information from large datasets, many things may need to be taken into account. How can we design intervention studies that take into account all these factors that will enable trustworthy conclusions to be drawn about the educational and information value of DV?

To illustrate the complexity of the factors and their associated interactions, in the first example we consider one of the tools published in 2012 by the UK Office for National Statistics. Figure 1 displays a static image of an interactive internal migration map 'frozen' for the migration into Cornwall. The tool can be found at
www.neighbourhood.statistics.gov.uk/HTMLDocs/dvc25/index.html.
We asked the constructor of this migration tool the following questions (answers in italics):

- the intended audience(s);
  'Originally, traditional users of the underlying dataset - local authorities, central government research organisations etc. By making the content engaging, we also wanted it to conceivably extend to a wider (but largely undefined audience) - design brief is to retain the complexity of the data without dumbing it down. So, within reason, this might therefore extend to educational uses, but it was not designed for that purpose'.

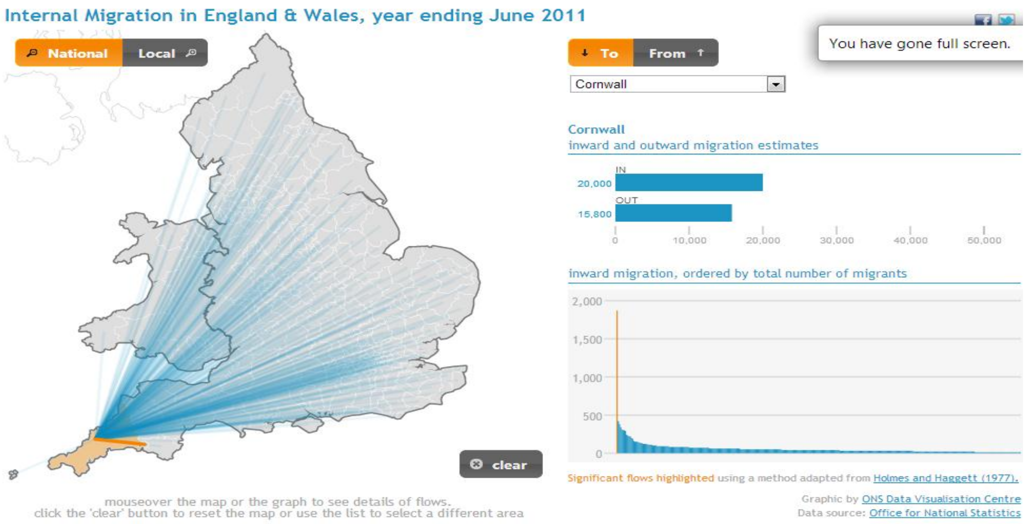

Figure 1. Static image of inward migration to Cornwall

- the intended key messages from using the tool;
  'It is an exploratory tool - we are not offering any pre-conceived stories - but we want people to find and share them. That's why we've worked to ensure that every time you interact with the map you get a customised URL'.
- what could or should be learned from using the map.
  'We have a general belief that if something is tactile and visual, it's more likely to grab the attention. We think you see things in the map that you don't see from looking at the underlying tables. We also think that the map is a quick way of getting to the underlying figures, using interaction with map or graph'.

As a further qualitative, focus group, start to evaluation we asked three people of known varying educational backgrounds the following questions (answers in italics):

- what they thought about the look and feel of the DV tool;
  Very nice, very slick and spot on; Need more control of the zoom; The selection of boundaries seems arbitrary; It would be good to have a grouping or a splitting option; The colours are not to my taste.
- the key messages that it communicated;
  There is no background information, so no key message other than people move; It would be helpful to have a pull-down menu with suggestions on using the tool; Key points should be made about the importance of census migration data – that they have an impact on transport policy, local planning, local services etc.
- what they could learn from using the DV tool.
  There are few clues as to what someone could learn from the map outside of professional usage - making it no more than an interesting curiosity; It may help people with genealogy research investigating the trends in migration; I was

interested in migration to Devon and Cornwall as it highlights the growth in second home ownership and migration to the southwest by others from outside the region.

We asked one experienced teacher what could be taught or learned from the interactive map and what age groups would get most from it. The response was:

- *The tool is easy and fun to use and children may be motivated by investigating their own areas; to be effective in the classroom it would need to be built into activities with careful guidance on how the information can be helpful.*

The comments from the small qualitatively-chosen sample of people showed, amongst other things, that they were motivated to use the DV to find out migration patterns to and from regions on the map they knew about or even used to live in. A question that follows is whether DV could be made more effective if users could see images of data within the tool that relate to their own experiences. Could this feature trigger a desire to learn more from the images, or could it even enable the user to retain more?

Further complexity arises from the extent and form of interaction allowed by the delivery software. In a second example, we consider the dynamic graphs popularized by Professor Hans Rosling (www.gapminder.org). They involve displaying circles that represent, in terms of their area and colour, summary statistics of countries. In Figure 2 we show a screen snap of a graph of life expectancy against income per person from 1810 to 2011, with India and the UK being the countries displayed.

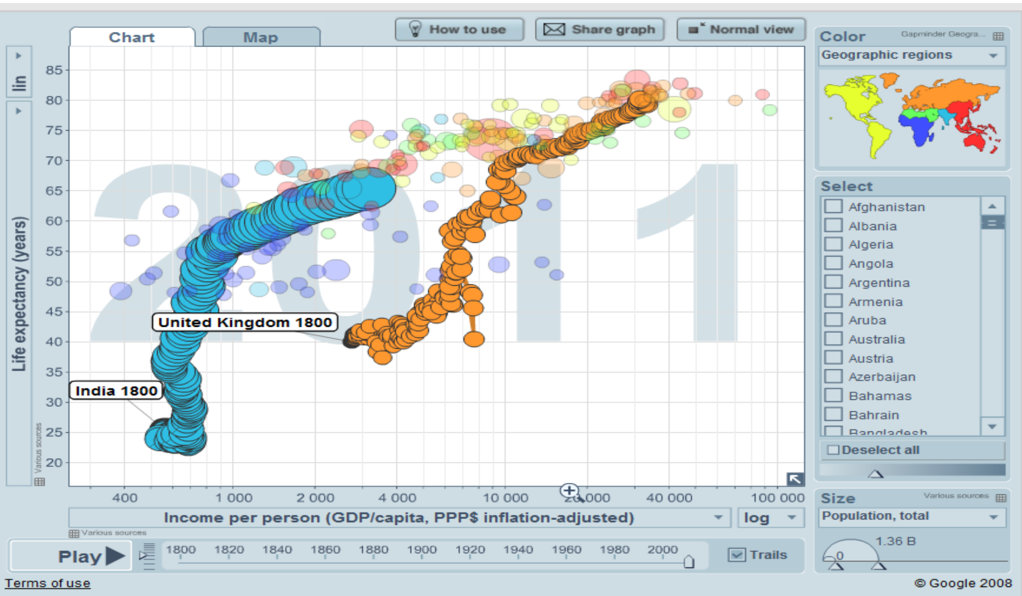

Figure 2. Static image of income per person for countries through time

The graph has no time axis; instead the overlapping circles represent the population size over time for both countries. After clicking the 'Play' button, the graph comes alive by displaying the build-up of the circles year by year - the graph is constructed gradually and dynamically. The dynamic process in Figure 1 is different; mouse-overs or clicks enable snapshots of migration patterns of information to be displayed on the map of England and Wales, with corresponding static charts displayed to the right. A question arises as to whether the *way* the DV appears can affect the educational value of the graphics, but restricted space in this paper prevents us from exploring this issue further.

There are many other DV tools, such as the open source software iNZight (www.stat.auckland.ac.nz/~wild/iNZight/) from the University of Auckland and the free software produced by the SMART Centre, Durham University (http://www.dur.ac.uk/smart.centre/projects/), which provide other innovative ways of displaying data. We do not have space to consider them here.

5. Conclusions

We have used the example of DV tools to illustrate the need to consider factors that might influence their effectiveness. In doing this we have shown it is important to use both qualitative and trustworthy quantitative approaches to assess their usefulness as statistical communication tools. These tools are used by both statisticians for communicating and informing and, as their use increases, they are likely to be used by statistical educators to assess their effectiveness in teaching and learning.

Pooling expertise from these two areas of statistics to advise running valid and efficient RCTs and taking account of the structure using multilevel statistical models to *measure* their effectiveness, could benefit statistics teaching in schools, higher education and in communicating to the public information contained in complex data sets.

Defining target populations is always important when working with data from which generalisations are to be made. However, in some published education research papers we have found many papers where this not done. The constructor of the tool that produced Figure 1 never intended it to be used in education, but presumably would be pleasantly surprised if it was effective. So in defining an *education* target population for a DV tool we may need to account for the fact that it is being used for a purpose it was never intended.

Also we have identified several factors connected with how this DV tool will be taught and learned and at least eight intervention-level factors that need to be considered for evaluating its effectiveness. Of course, it may be that some of these factors can be eliminated either through good design and RCTs, or allowed for by building them into a nested data response structure within a multilevel model. The ASA report gave three examples where RCTs are used in educational interventions. However, to our surprise we have not been able to identify the use of RCTs or multilevel modelling in any published *statistical* education interventions.

We feel that it is important that statisticians who are experts in these areas should work together to encourage the use of good statistical practice for assessing the effectiveness of interventions for improving statistics pedagogy, including emerging DV tools. The methods in this paper will form part of a research project ‘Communicating Statistics and Data Visualisation’ that will evaluate the effectiveness of a variety of DV tools in education and for the general public. It will involve statisticians and statistical educators and will be carried out at Plymouth University in the next three years.

6. References

Goldacre, B. (2013). *Building evidence into education*. Available at www.education.gov.uk/inthenews/inthenews/a00222740/building-evidence-into-education.

Goldstein, H. (2010). *Multilevel statistical models (4th edition)*. Oxford: Wiley.

Hutchison, D. & Styles, B. (2010). *A guide to running randomised controlled trials for educational researchers*. Slough: NFER.